Kinesis

Thursday, July 9, 2020 11:50 PM

- Stream needs manual provision, Firehouse scale automatically

Dynamo DB

Friday, July 10, 2020 12:28 AM

1 WCU = 1KB

1 RCU = 4KB (eventual) or 4KB /2 (strong consistent)

GetItem -> PrimaryKey (BatchGetItem = more tables)
Query -> sort through data for the retrieve
Filter -> After retrieving
Scan -> Entire table

DynamoDB Streams captures a time-ordered sequence of item-level modifications in any DynamoDB table and stores this information in a log for up to 24 hours.

KEYS_ONLY — Only the key attributes of the modified item.

NEW_IMAGE — The entire item, as it appears after it was modified.

OLD_IMAGE — The entire item, as it appeared before it was modified.

NEW_AND_OLD_IMAGES — Both the new and the old images of the item.

DynamoDB has a feature called “Conditional Update / Delete”
• That means that you can ensure an item hasn’t changed before altering it
• That makes DynamoDB an optimistic locking / concurrency database

Global Secondary Index
Used to speed up queries on non-key attributes.

Local Secondary Index
Provides an alternative sort key to use for scans and queries.

Amazon DynamoDB – Performance and Throttling

- Reduce the frequency of requests and use exponential backoff.
- Try to design your application for uniform activity across all logical partition keys in the table and its secondary indexes.
- Use burst capacity effectively - DynamoDB currently retains up to 5 minutes (300 seconds) of unused read and write capacity which can be consumed quickly

DAX is a managed service that provides in-memory acceleration for DynamoDB tables.
Ø Provides managed cache invalidation, data population, and cluster management.
Ø DAX is used to improve READ performance (not writes)

SQS

Friday, July 10, 2020 12:31 AM

Default visibility time out = 30 sec. API = ChangeVisibilityTimeOut

Fan Out = SNS + a number of SQS

SQS does NOT same-time deliver to multi consumers

SQS max size message = 256 KB, Use Extendted Client side library up to 2 GB

FIFO queue with dedup id - 创造 5 min的 dedup interval for the same message

API:

1. WaitTimeSeconds >0 is long polling
2. ReceiveMessage - retrieves one or messages up to 10

Delay Queue:
Postpones delivery of new messages to a queue for a number of seconds.
Messages sent to the Delay Queue remain invisible to consumers for the duration of the delay period.

Step Functions

Friday, July 10, 2020 12:46 AM

Written in Json

built on Lambda and is and orchestration service that lets you easily coordinate multiple Lambda functions into flexible workflows that are easy to debug and easy to change.

CloudFormation

Friday, July 10, 2020 12:46 AM

- Resources (Must), Parameters, Conditions, Outputs
- Only pay for underlying resources - Same with Beanstalk
  CF = infra as code. BeanStalk = no infra to manage, PaaS

StackSets extends the functionality of stacks by enabling you to create, update, or delete stacks across multiple accounts and regions with a single operation.

Change sets allow you to preview how proposed changes to a stack might impact your running resources.

Nested stacks allow re-use of CloudFormation code for common use cases.

ECS

Sunday, July 12, 2020 1:51 PM

Fargate: The Fargate launch type allows you to run your containerized applications without the need to provision and manage the backend infrastructure.

ECR - Amazon Elastic Container Registry (ECR) is a fully-managed Docker container registry that makes it easy for developers to store, manage, and deploy Docker container images.

You need to run some commands to push pull:
- $(aws ecr get-login --no-include-email --region eu-west-1) / aws ec2 get-login-password
- docker push 1234567890.dkr.ecr.eu-west-1.amazonaws.com/demo:latest
- docker pull 1234567890.dkr.ecr.eu-west-1.amazonaws.com/demo:latest

A *task placement strategy* is an algorithm for selecting instances for task placement or tasks for termination. Task placement strategies can be specified when either running a task or creating a new service.

binpack - place tasks based on the least available amount of CPU or memory. This minimizes the number of instances in use.

random - place tasks randomly.

spread - place tasks evenly based on the specified value. Accepted values are instanceId (or host, which has the same effect), or any platform or custom attribute that is applied to a container instance, such as attribute:ecs.availability-zone. Service tasks are spread based on the tasks from that service. Standalone tasks are spread based on the tasks from the same task group.

S3

Sunday, July 12, 2020 2:13 PM

Single file = 5 TB, bucket no limit
Multi-part upload = 5GB

Principle - allow or deny to a policy

Variable - ${aws: username} generalized condition as a placeholder. 此处=任何 current user

Object Key: · Development/Projects.xls
The console uses the key name prefixes (Development/, Finance/, and Private/) and delimiter ('/') to present a folder structure as shown.

Client-side encryption is the act of encrypting data **before (server side encryption afterwards**_)sending it to Amazon S3. You have the following options:

- Use a customer master key (CMK) stored in AWS Key Management Service (AWS KMS).
- Use a master key you store within your application.
- Additionally, using HTTPS/SSL to encrypt the data as it is transmitted over the Internet adds an additional layer of protection.

Server-side encrytion:
1.S3 managed keys: fully managed
2.KMS-manage keys: Customer master keys 1) added protection as separate permission 2) audit trail

	Who en/decrypts the data	Who stores the secret	Who manages the secret
SSE-AES	AWS	AWS	AWS
SSE-KMS (AWS managed CMK)	AWS	AWS	AWS
SSE-KMS (customer managed CMK)	AWS	AWS	you
SSE-C	AWS	you	you

Amazon S3 Encryption

client-side

Previou

Encryption Option	How it Works
SSE-S3	Use S3's existing encryption key for AES-256
SSE-C	Upload your own AES-256 encryption key which S3 uses when it writes objects
SSE-KMS	Use a key generated and managed by AWS KMS
Client-Side	Encrypt objects using your own local encryption process before uploading to S3

CloudFront

Sunday, July 12, 2020 7:05 PM

Global Service: CloudFront, IAM

CI/CD

Sunday, July 12, 2020 7:15 PM

CodePipeline:
Each stage contains at least one action.
Artifacts are passed stored in Amazon S3 and passed on to the next Stage

CodeDeploy : to EC2 and Lambda functions

CodeBuild: Build instructions can be defined in code (buildspec.yml file)
Buildspec.yml - build instructions
Appspec.yml - deployment actions

CodeStar: Each AWS CodeStar project comes with a project management dashboard, including an integrated issue tracking capability powered by Atlassian JIRA Software.

Beanstalk

Sunday, July 12, 2020 7:27 PM

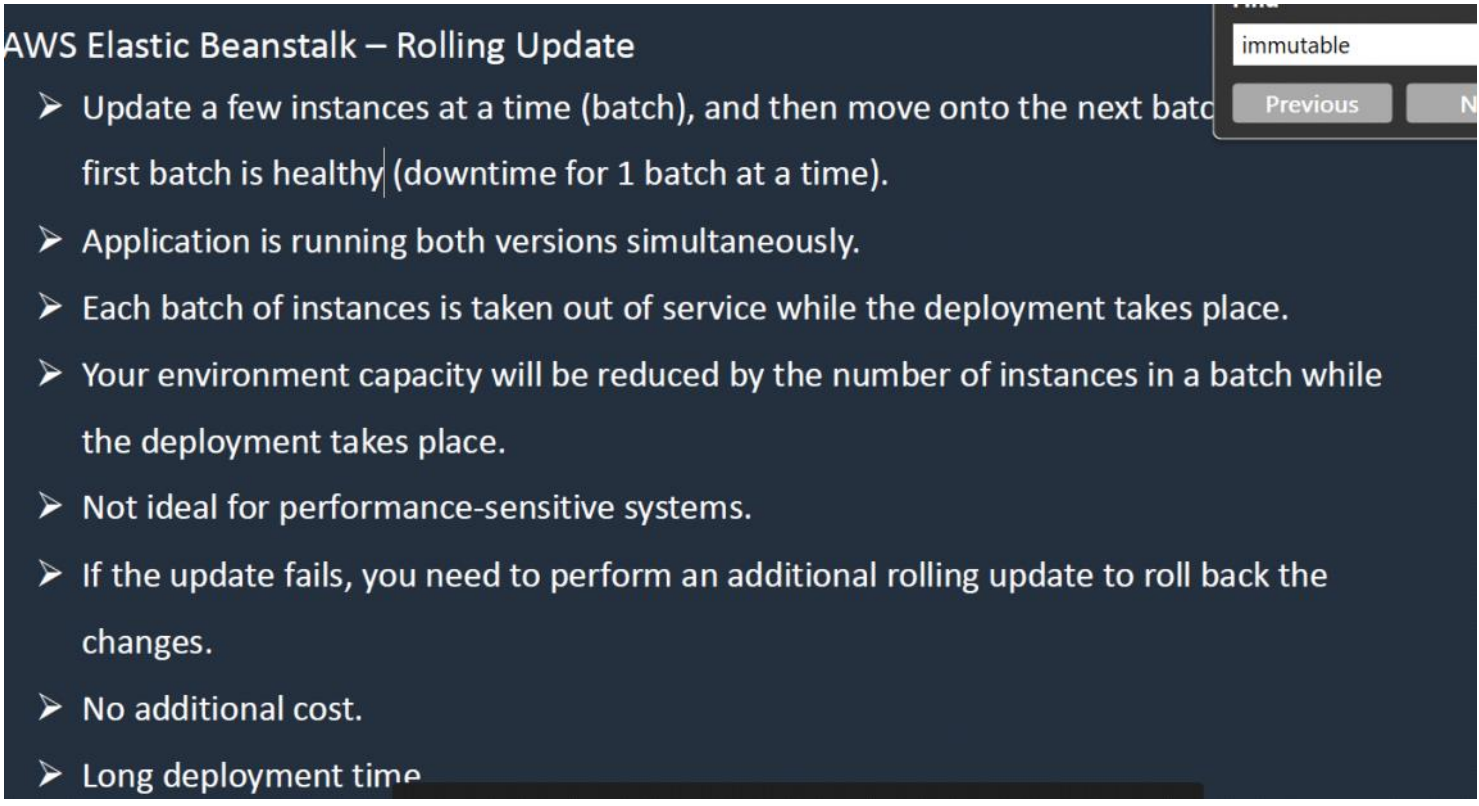
Each environment contains one and only one application version

Config extensions:
in the .ebextensions/ directory in the root of source code
config extensions (example: logging.config)

Build Docker env:

Define the containers in the Dockerrun.aws.json file in YAML format and save at the root of the source directory" is incorrect because the contents of the file should be in JSON forma

BeanStalk Deployment:
All at once, Rolling, Rolling with additional batches, immutable, Blue/Green (Use Route 53 so not a native feature)



AWS Elastic Beanstalk – Rolling with Additional Batch Update

- Like Rolling but launches new instances in a batch ensuring that there is full availability.
- Application is running at capacity.
- Can set the batch size.
- Application is running both versions simultaneously.
- Small additional cost.
- Additional batch is removed at the end of the deployment.
- Longer deployment.
- Good for production environments.

AWS Elastic Beanstalk – Immutable Update

- Launches new instances in a new ASG and deploys the version update to these instances before swapping traffic to these instances once healthy.
- Zero downtime.
- New code is deployed to new instances using an ASG.
- High cost as double the number of instances running during updates.
- Longest deployment.
- Quick rollback in case of failures.
- Great for production environments.

API Gateway deployment:
Canary (This is blue / green deployment with AWS Lambda)

Lambda Deployment:

1. We can define a “dev”, ”test”,“prod” aliases and have them point at different lambda versions

• Aliases are mutable (versions are not, except the $lastest)
• Aliases enable Blue / Green deployment by assigning weights to lambda functions

2. Use CodeDeploy for Lambda: linear, all-at-once, Canary

Lambda

Sunday, July 12, 2020 7:28 PM

Version error - you can not do anything but wait for new release

Event Mappings

Services That Lambda Reads Events From

Amazon Kinesis

Amazon DynamoDB

Amazon Simple Queue Service

Limits

You can use the /tmp directory if the function needs to download a large file or disk space for operations. The maximum size is 512 MB.

Memory allocation 128MB - 3008MB in 64MB increments.
Maximum execution time is 15 minutes (900 seconds).

Lambda function deployment size is 50 MB (zipped), 250 MB unzipped.

Burst Concurrency Limits:
Ø 3000 – US West (Oregon), US East (N. Virginia), Europe (Ireland).
Ø 1000 – Asia Pacific (Tokyo), Europe (Frankfurt).
Ø 500 – Other Regions.
After the initial burst, your functions' concurrency can scale by an additional 500 instances each minute. This continues until the account limit (default 1000 exec/sec is reached).

ELB

Sunday, July 12, 2020 7:37 PM

Application load balancers (Layer 7) -ALB support HTTP/HTTPS & Websockets protocols

Network load balancers (Layer 4) allow to do: • Forward TCP traffic to your instances • Handle millions of request per seconds (less latency only)

X forwarded for header = identify IP address

Cognito

Sunday, July 12, 2020 10:29 PM

User Pool - users can sign in to a web or mobile app through Amazon Cognito COGNITO_USER_POOLS authorizer

Identiy Pool - With an identity, you can obtain temporary, limited-privilege AWS credentials to access other AWS services.

Exam tip: To make it easier to remember the difference between User Pools and Identity Pools, think of Users Pools as being similar to IAM Users (Authenticate) or Active Directory and Identity Pools as being similar to an IAM Role (authorize)

Web Identify Federation
AWS Cognito works with external identity providers that support SAML or OpenID Connect, social identity providers (such as Facebook, Twitter, Amazon)

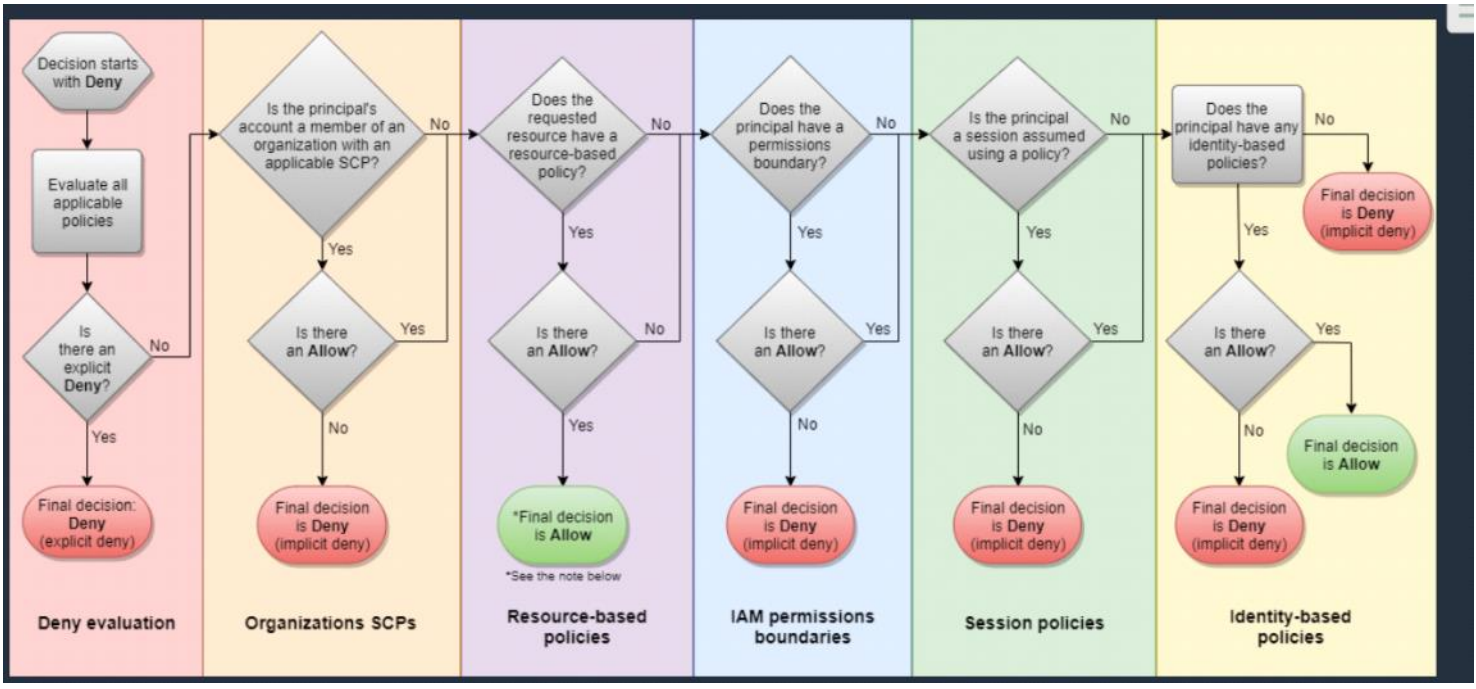
IAM

Wednesday, July 15, 2020 12:10 AM

IAM Policy Evaluation Logic

Identity-based policies – Identity-based policies are attached to an IAM identity (user, group of users, or role) and grant permissions to IAM entities (users and roles).
Ø Resource-based policies – Resource-based policies grant permissions to the principal (account, user, role, or federated user) specified as the principal.
Ø IAM permissions boundaries – Permissions boundaries are an advanced feature that sets the maximum permissions that an identity-based policy can grant to an IAM entity (user or role).
Ø AWS Organizations service control policies (SCPs) – Organizations SCPs specify the maximum permissions for an organization or organizational unit (OU).
Ø Session policies – Session policies are advanced policies that you pass as parameters when you programmatically create a temporary session for a role or federated user.

Ø By default, all requests are implicitly denied. (Alternatively, by default, the AWS account root user has full access.)
Ø An explicit allow in an identity-based or resource-based policy overrides this default.
Ø If a permissions boundary, Organizations SCP, or session policy is present, it might override the allow with an implicit deny.
Ø An explicit deny in any policy overrides any allows.



EC2 Policy:

The customer-managed policy is more secure in this situation as it can be locked down with more granularity to ensure the EC2 instances can only read and write to the specific bucket.

With an AWS managed policy you must choose from read only or full access and full access would provide more access than is required:

API Gateway

Wednesday, July 15, 2020 3:22 PM

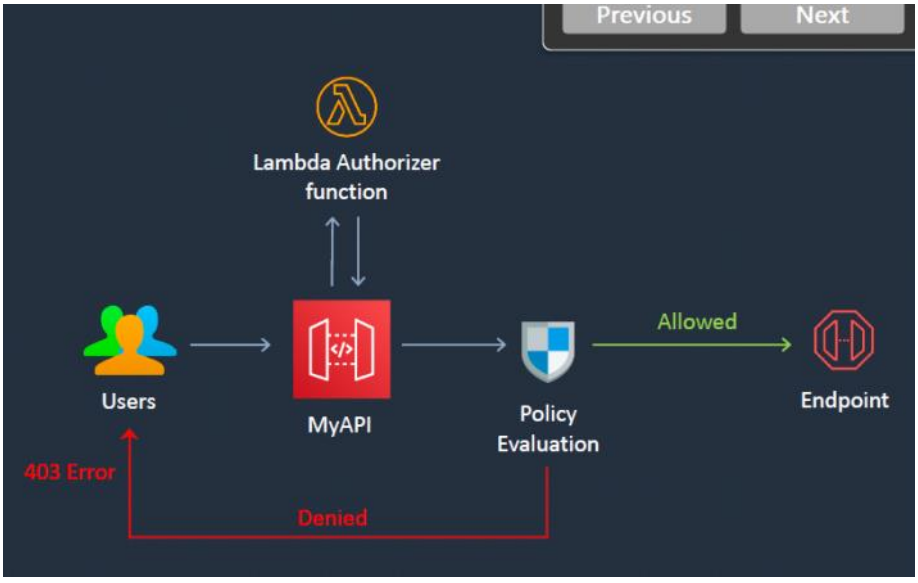
Caching is enabled for a stage.

- When you enable caching for a stage, API Gateway caches responses from your endpoint for a specified time-to-live (TTL) period, in seconds.
- The default TTL value for API caching is 300 seconds. The maximum TTL value is 3600 seconds. TTL=0 means caching is disabled.
- Invalidate an existing cache entry and reload it from the integration endpoint for individual requests. The request must contain the `Cache-Control: max-age=0` header

A *Lambda authorizer* (formerly known as a *custom authorizer*) is an API Gateway feature that uses a Lambda function to control access to your API.

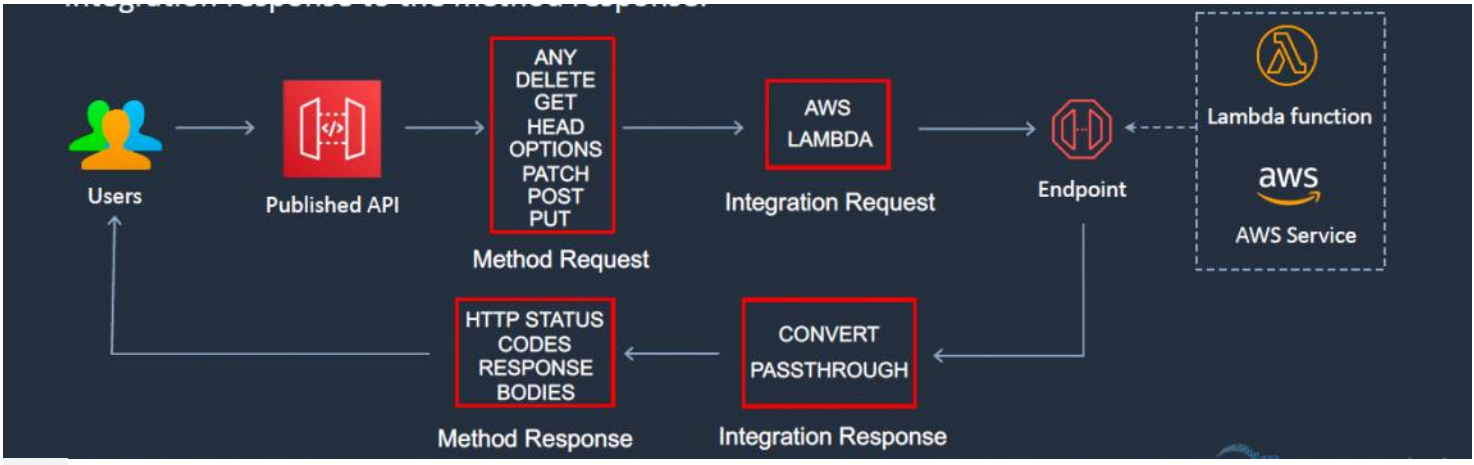
- A *token-based* Lambda authorizer (also called a TOKEN authorizer) receives the caller's identity in a bearer token, such as a JSON Web Token (JWT) or an OAuth token.
- A *request parameter-based* Lambda authorizer (also called a REQUEST authorizer) receives the caller's identity in a combination of headers, query string parameters, stageVariables, and $context variables.

Ø For WebSocket APIs, only request parameter-based authorizers are supported.

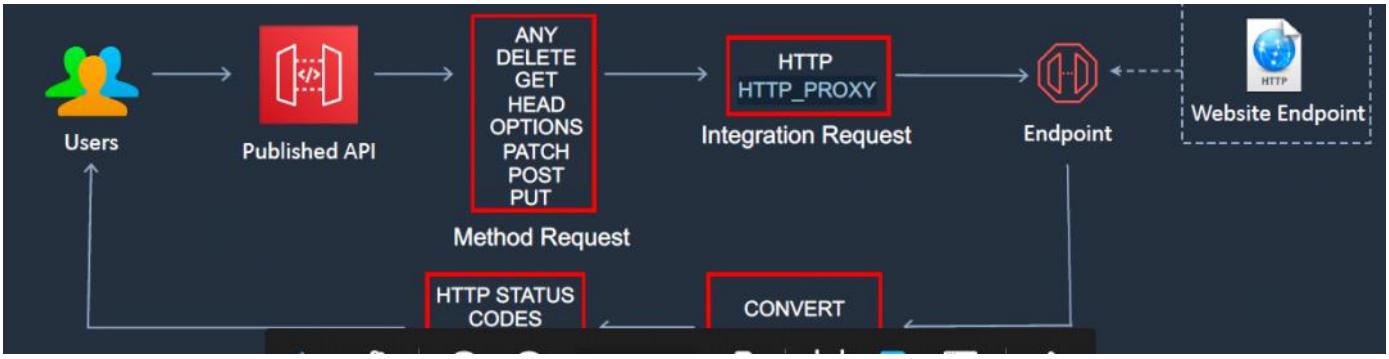


Integration Types:

AWS



HTTP



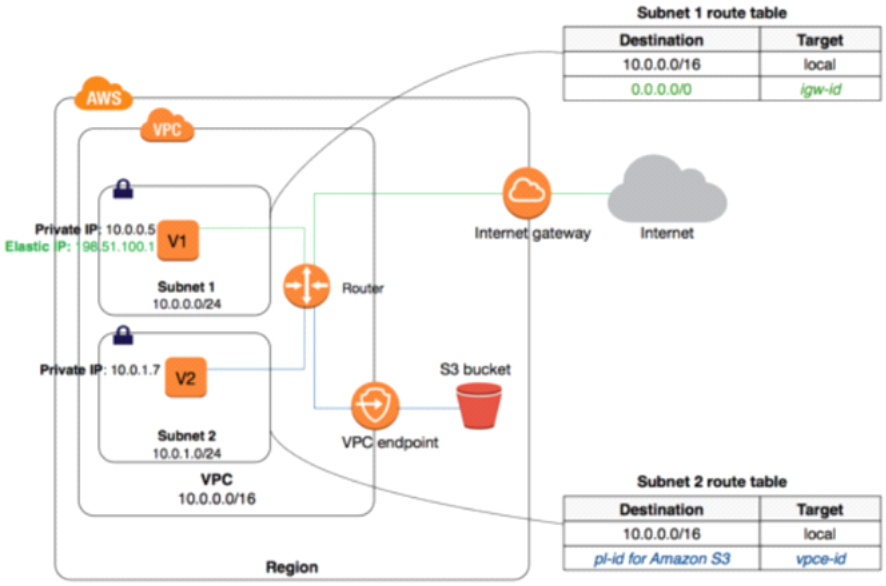
AWS proxy /HTTP proxy : no need to set request and response

VPC

Wednesday, July 15, 2020 3:35 PM

EC2 Internet Connectivity:

The first option is to enable Internet connectivity through either a NAT Gateway or a NAT Instance. The specific type of VPC endpoint to S3 is a Gateway Endpoint. EC2 instances running in private subnets of a VPC can use the endpoint to enable controlled access to S3 buckets, objects, and API functions that are in the same region as the VPC.



Security

Wednesday, July 15, 2020 4:23 PM

AWS Systems Manager Parameter Store（区别secrets manager，后者自动rotate password)

- provides secure, hierarchical storage for configuration data management and secrets management. You can store data such as passwords, database strings, and license codes as parameter values.
- A secure string parameter is any sensitive data that needs to be stored and referenced in a secure manner.
- You can store data such as passwords, database strings, and license codes as parameter values. No native rotation of keys (difference with Secrets Manager which does it automatically; KMS has automatic rotation which is the backend service for SSM parameter store).

AWS CloudHSM
is a cloud-based hardware security module (HSM) that enables you to easily generate and use your own encryption keys on the AWS Cloud (区别 KMS)

AWS Web Application Firewall (WAF)
helps protect web applications from attacks by allowing you to configure rules that allow, block, or monitor (count) web requests based on conditions that you define.

IAM roles are allow /deny policies. While S3 bucket policy sticks to a bucket

S3 encrypt - http header x-amz-server-side-encryption

IAM policy - json must have: resource, effect, action

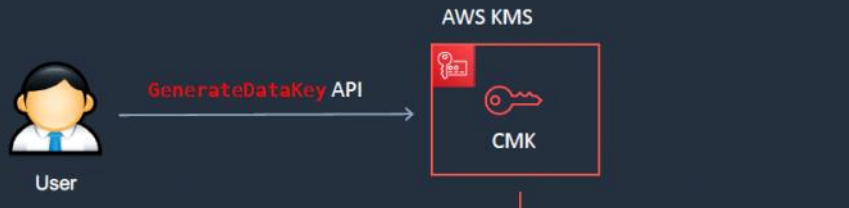
KMS: (区别S3 client-side encryption)
Customer master keys are the primary resources in AWS KMS. The CMK includes metadata, such as the key ID, creation date, description, and key state.

Type of CMK	Can view	Can manage	Used only for my AWS account	Automatic rotation
Customer managed CMK	Yes	Yes	Yes	Optional. Every 365 days
AWS managed CMK	Yes	No	Yes	Required. Every 1095 days
AWS owned CMK	No	No	No	Varies

AWS KMS API and CLI

- Encrypt (aws kms encrypt):
  - Encrypts plaintext into ciphertext by using a customer master key
  - You can encrypt small amounts of arbitrary data, such as a persona password, or other sensitive information.
  - You can use the Encrypt operation to move encrypted data from o
- Decrypt (aws kms decrypt):
- Decrypts ciphertext that was encrypted by an AWS KMS customer mas the following operations:
  - Encrypt
  - GenerateDataKey
  - GenerateDataKeyPair
  - GenerateDataKeyWithoutPlaintext
  - GenerateDataKeyPairWithoutPlaintext

AWS KMS – Data Encryption Keys



- Data keys are encryption keys that you can use to encrypt data, including large amounts of data and other data encryption keys.
- You can use AWS KMS customer master keys (CMKs) to generate, encrypt, and decrypt data keys.
- AWS KMS does not store, manage, or track your data keys, or perform cryptographic operations with data keys.
- You must use and manage data keys outside of AWS KMS.

X-ray

Wednesday, July 15, 2020 5:44 PM

Annotations - Use annotations to record information on segments or subsegments that you want indexed for search.

Metadata - Key / value pairs, not indexed and not used for searching.

X-Ray SDK is installed in your application and forwards to the X-Ray daemon which forwards to the X-Ray API. (Daemon on the other hand can be used for EC2, ecs, beanstalk, lambda)

STS

Wednesday, July 15, 2020 11:40 PM

The AWS Security Token Service (STS) is a web service that enables you to request temporary, limited-privilege credentials for IAM users or for users that you authenticate (federated users)

Also has Web Identity Federation

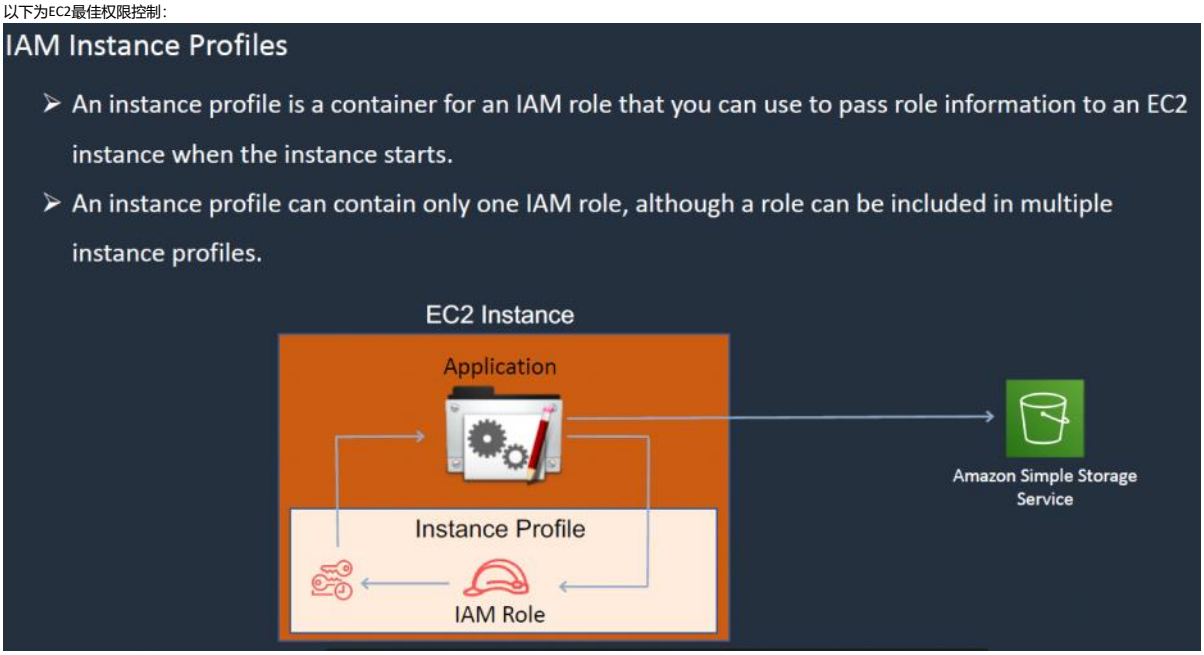
EC2 Auto Scaling – Scaling Option

Scaling Option	What it is	When to use
Maintain	Ensures the required number of instances are running	Use when you always need a known number of instances running at all times
Manual	Manually change desired capacity via the console or CLI	Use when your needs change rarely enough that you're OK to make manual changes
Scheduled	Adjust min/max instances on specific dates/times or recurring time periods	Use when you know when your busy and quiet times are. Useful for ensuring enough instances are available *before* very busy times
Dynamic	Scale in response to system load or other triggers using metrics	Useful for changing capacity based on system utilization, e.g. CPU hits 80%

EC2 Auto Scaling – Scaling Types (associated with Dynamic Scaling Policies)

Scaling	What it is	When to use
Target Tracking Policy	The scaling policy adds or removes capacity as required to keep the metric at, or close to, the specified target value	A use case is that you want to keep the aggregate CPU usage of your ASG at 70%
Simple Scaling Policy	Waits until health check and cool down period expires before re-evaluating	This is a more conservative way to add/remove instances. Useful when load is erratic. AWS recommend step scaling instead of simple in most cases
Step Scaling Policy	Increase or decrease the current capacity of your Auto Scaling group based on a set of scaling adjustments, known as step	Useful when you want to vary adjustments based on the size of the alarm breach

以下为EC2最佳权限控制：



Cloud Watch

Monday, July 20, 2020 2:32 PM

Amazon CloudWatch Logs Agent

- The CloudWatch Logs agent provides an automated way to send log data to CloudWatch Logs from Amazon EC2 instances.
- There is now a unified CloudWatch agent that collects both logs and metrics.
- The unified CloudWatch agent includes metrics such as memory and disk utilization